Proceeding of the Electrical Engineering Computer Science and Informatics
Vol. 7, October 2020 **ISSN: 2407-439X**

Combination of Genetic Algorithm and Brill Tagger Algorithm for Part of Speech Tagging Bahasa Madura

Nindian Puspa Dewi
Departement of Informatics
Universitas Madura
Pamekasan, Jawa Timur, Indonesia
nindianpd@unira.ac.id

Ubaidi
Departement of Informatics
Universitas Madura
Pamekasan, Jawa Timur, Indonesia
ubaidi@unira.ac.id

Joan Santoso
Departemen of Information Technology
Institut Sains dan Teknologi Terpadu Surabaya
Surabaya, Jawa Timur, Indonesia
joan@stts.edu

Eka Rahayu Setyaningsih
Departemen of Information Technology
Institut Sains dan Teknologi Terpadu Surabaya
Surabaya, Jawa Timur, Indonesia
eka@stts.edu

***Abstract*—Part of speech (POS) is commonly known as word types in a sentence such as verbs, adjectives, nouns, and so on. Part of Speech (POS) Tagging is a process of marking the word class or part of speech in every word in a sentence. Part of Speech Tagging has an important role to be used as a basis for research in Natural Language Processing. That is why research on Part of Speech Tagging for Bahasa Madura as an effort to preserve and develop the use of regional languages. In this research, POS Tagging is done using the Brill Tagger Algorithm which is combined with the Genetic Algorithm. Brill Tagger is a POS Tagging Algorithm that has the best level of accuracy when implemented in other languages. Genetic Algorithms used in the contextual learner process with consideration in previous studies can increase the speed of the training process so that it is more efficient. The results of this study are then compared with the results of the previous study so that we can find out suitable algorithms used for the development of text processing in Bahasa Madura. From a series of experiments, the average accuracy obtained by using Brill Tagger is 86.4% with the highest accuracy of 86.7%, while using GA Brill Tagger shows an average accuracy of 86.5% with the highest accuracy of 86.6%. Testing by observing OOV (Out of Vocabulary) achieves an average accuracy of 67.7% for Brill Taggers and 64.6% for GA Brill Taggers. Testing by considering multiple POS with Brill Tagger produces an average accuracy of 73.3% while testing using GA Brill Tagger produces an average accuracy of 90.9%. This shows that the accuracy with GA Brill Tagger is better than Brill Tagger, especially if considering multiple POS. This is because GA Brill Tagger can generate rules for handling the existence of multiple POS more than pure Brill Tagger.**

Keywords*—*pos tagging, brill tagger, genetic algorithm, bahasa madura

I. INTRODUCTION

Today, technological advances have learned about human language. Many studies have been conducted to process natural language into a computational model. This allows interaction between humans and computers to occur using human language (natural language). Research in this field became known as Natural Language Processing. One study in Natural Language Processing is Part-of-Speech Tagging.

Part-of-Speech (POS) is known as word types in a sentence [1] such as verbs, adjectives, nouns, etc. Part-of-Speech (POS) tagging is a process of marking the word class for each word in a sentence. POS Tagging is a basis of research in Natural Language Processing, such as in Word Sense Disambiguation, Stemming in Information Retrieval, and Question and Answering [2].

Research on Part-of-Speech Tagging in Indonesia has been carried out using various methods including POS Indonesian Tagging with Hidden Markov Model and Rule Based [3], Probabilistic Part of Speech Tagging for Indonesian [4] using 37 tag set, Brill Tagger Implementation to provide POS Tagging on Indonesian Language Documents [5], Toward a Standardized and More Accurate Indonesian Part-of-Speech Tagging [6], and On Part of Speech Tagger for Indonesian Language [7]. From several studies that have been done, the highest accuracy value is by using Brill Tagger [8]. Brill Tagger was first introduced by [9]. The Tagger process is a transformation or rules of learning outcomes from detecting error values [10]. From several studies on POS Tagging, the highest accuracy value is to use the Brill Tagger method. Brill Tagger has also applied in many languages, such as English, Kadazan, and Bahasa Indonesia.

POS Tagging research using genetic algorithms such as Part-of-Speech Tagging using Genetic Algorithms [11], A New Approach to the POS Tagging Problem Using Evolutionary Computation [12], and Genetic Algorithm (GA) Implementation for Feature Selection in POS Tagging Manipuri [13]. Research that combines Brill Tagger and Genetic Algorithm, was carried out by Wilson who included GA in Brill Tagger to improve time efficiency compared to using Brill Tagger alone [14]. Another study, Genetic Algorithms in the Brill Tagger written by Johannes Bjerva, explained that Brill GA-Tagger performed much better than standard Brill tagger in all 9 target languages [15].

Bahasa Madura is a regional language used by ethnic Madurese, both living on Madura Island and outside the island, as a means of daily communication. The area of Bahasa Madura usage is not only limited to Madura Island but also extends to other places outside the island such as Sapudi, Raas, Goat, Kangean, and other surrounding islands because the majority of the islands are inhabited by Bahasa Madura. Bahasa Madura as a regional language needs to be fostered and developed, especially as a means of developing regional culture and national culture [16].

In previous studies, we have conducted POS Tagging research in Bahasa Madura using the Brill Tagger Algorithm [17], [18]. In this study, we used Brill Tagger combined with genetic algorithms (GA Brill Tagger). The difference with previous research, besides using GA Brill Tagger for POS Tagging in Bahasa Madura, this research also conducted experiments using words that have multiple POS. Multiple POS means words that have more than one class of words or tagset, such as the word "bisa" in Indonesian that can have tagset modals (MD) and tagset Noun (NN). The results of this study are then compared with the results of the previous study so that we can find out suitable algorithms used for the development of text processing in Bahasa Madura.

II. METHODOLOGY

A. Brill Tagger

Brill Tagger introduced by Eric Brill in 1992. Generally, Brill Tagger is also called Transformation-based Error-driven Learning (TEL). Brill Taggers are the basis of transformation or rules and learn from detecting error values [9].

Brill Tagger can give the right word class to a word by using lexical and contextual rules. Lexical rules are the result of lexical learners. Lexical rules are rules used to label words based on word affixes. Contextual rules are rules that pay attention to the existence of tags around the word being checked or searched for labels [19]. Contextual rules are the result of contextual learners.

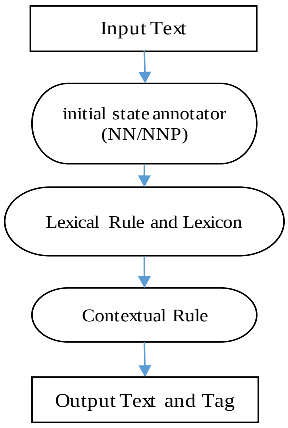

Fig. 1. Brill tagger.

The labeling process in Brill Tagger starts with the initial tagging process. This process is usually called the initial state annotator and involves the lexicon file. The next process is the formation of Final State Tagger and involves a lexical or a contextual rule file, as in Figure 1.

The initial input from Brill Tagger is called Unannotated Word. It is a text file that contains words that not labeled. Every word in the text will be given an initial word-class through the initial state annotator process. Initial state annotator can be done in two ways, namely giving initial NN (Common Noun) tags to all words or by giving NNP (Proper Common Noun) tags for words those beginning with capital letters and NN for those not preceded by capital letters. The results of labeling from this process then referred to as Temporary Corpus.

The next step is to check with the lexical rule. Each rule in lexical rules applied to words. The results of this process are called pretagged. The words that do not have a tag, it will be labeled according to the lexicon. The next step is to check with the contextual rule. Input at this process is the output of the Start State Tagger stage. Each rule at contextual rules applied to words that are not in the lexicon.

B. Genetic Algorithm

Genetic algorithm is a search method based on the natural evolutionary process [20], namely the formation of a random initial population consisting of individuals with traits that depend on genes on their chromosomes. Individuals carry out reproductive processes to give birth to offspring. Offspring formed from a combination of the properties of the two parents.

Like natural processes that inspire computational processes, populations in Genetic Algorithms also consists of many individuals called chromosomes. If in natural processes chromosomes contain unique individual characteristics, then in the Genetic Algorithm, chromosomes are representations of problem solving that are still symbolic.

As with the natural selection process, only fit individuals survive in the population. Each generation, chromosomes will undergo an evaluation process using the fitness function. The fitness value of a chromosome shows the quality of a chromosome in the population. The higher the fitness value of a chromosome, the higher the possibility to be maintained in the next population.

The initial chromosomes formed randomly and referred to as the parent. The chromosomes created from the parent chromosome pair are called child (offspring). The process of making a child from its parent is called a crossover operator. This process allows the child to inherit the properties of both parents [21]. In genetic algorithms, there is also a mutation operator (mutations). It is a process that can change genes in a chromosome.

The first time, a cycle of genetic algorithms developed by David Goldberg [22]. An overview of these cycles shown in Figure 2 below.

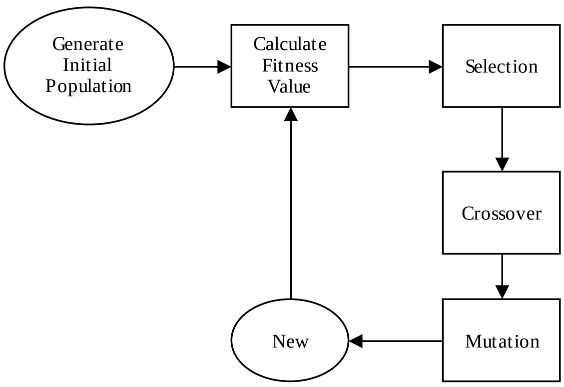

Fig. 2. Genetic algorithm.

C. *GA Brill Tagger*

Brill Tagger is an algorithm that has the best level of accuracy when implemented in English, Indonesian, and several other languages. In this study, Brill Tagger combined with Genetic Algorithms used in the training process, namely in the contextual learner section.

POS Tagging process in this study, starting from the preparation of data sets, the training process to testing. The process of creating a dataset begins with determining a standard tagset for Bahasa Madura. The next process is gathering a number of words in Madurese. These data are then extracted and referred to as unannotated text. Based on the predetermined tagset, this unannotated text is then tagged manually, so it becomes annotated text (goal corpus).

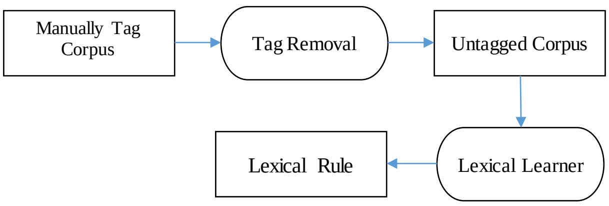

Fig. 3. Lexical learner.

Figure 3 explains the training process in a lexical learner [2]. In the training process, Manually Tag Corpus is needed which is a corpus that is tagged manually. Manually the corpus tag is then removed and called untagged corpus. Untagged corpus then compared with the corpus tag manual according to the lexical rule template.

The Lexical Learner will check tags based on affixes and suffixes on words. In this study, the contextual learner process uses genetic algorithms. The interpretation of the rules in an individual is shown below in figure 4 and 5.

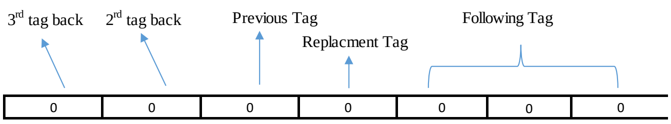

Fig. 4. Interpretation of rule in GA brill tagger.

The 3rd tag back shows the 3 tags before the correct tag (replacement tag). The 2nd back tag shows 2 tags before the replacement tag, and the previous tag shows the previous tag. Whereas, the following tags are the tags that follow the replacement tags. Each tag presented in 6 bits encoded in decimal according to the tagset sequence number. Following is an example of a line chromosome interpretation consisting of 3 rules.

0	0	9 (PRP)	1(VBT)	5(NN)	0	0

16 (IN)	3 (ADJ)	0	2 (VBI)	10 (NN)	32 (:;)	7 (NN)

0	0	9 (NN)	2 (RB)	29 (JJ)	0	0

Fig. 5. Example of rule interpretation in GA brill tagger.

Rule 1 shows change the tag to VBT if the previous tag is PRP and the next tag is NN. Rule 2 means to change the tag to VBI if the 3 previous tags are IN, the 2 previous tags are ADJ, the next tags are PTT, the 2 tags after that are :; and the 3rd tag that follows is NN.

Fitness value is the accuracy of each rule/individual compared to the goal corpus. The higher the fitness value of an individual, the higher the probability that the rule is correct.

$$Fitness = \frac{CR}{NCR} X \frac{CR}{TGC} \tag{1}$$

Where CR is number of corpus that matches the rule, NCR is number of corpus that matches the rule but ignores replacemant tags, and TGC is total of corpus sentences compared.

Figure 6 explains the Contextual Learner with GA. The randomly generated initial population (rules) are applied to the corpus dummy which is then compared to manually tagged corpus. After that, the fitness calculation for each chromosome is performed. Then each population is extracted according to the template and the fitness value is searched. Extractions that have greater fitness equal to the threshold will be stored in contextual rules. Individual selection is carried out with a tournament selection, where a group of randomly selected individuals will be journalized and then the two best individuals will be taken as parents. Then the crossovers and mutations are carried out to produce new individuals (new rules). The above steps continue to be repeated until a certain iteration or if no more chromosomes are found with better fitness values.

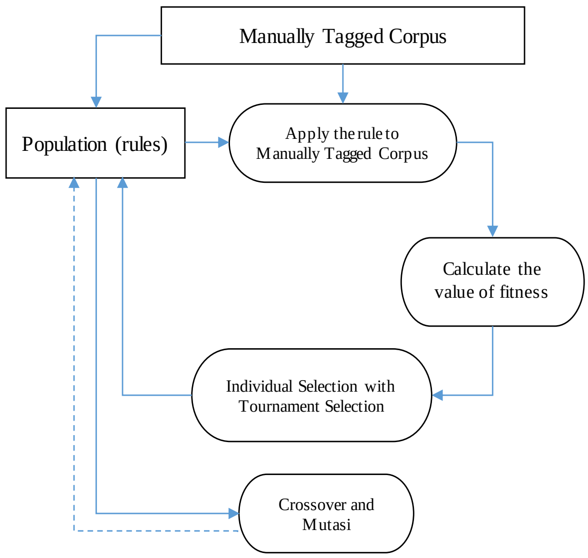

Fig. 6. Contextual learner with GA.

III. EXPERIMENT

A. *Data*

The experiment was carried out using a tagset consisting of 34 tagset as in [17]. The compilation of datasets was carried out by collecting articles of Bahasa Madura totaling 10,535 words [18] and manually tagged using a tagset. The results of this labeling are then referred to as Manually Tagged copus (Goal Corpus).

The structure of Bahasa Madura is almost the same as Bahasa Indonesia, so the determination of the word class is also not much different. It's just that there are a number of word classes broken down as if in Bahasa Indonesia [23], verbs are simply given a verb word class (VB), then in this study, it is divided into transitive verbs (VBT) and intransitive verbs (VBI).

B. *Result*

We conducted experiments using computers with the specifications of Intel Corei5 1.7 GHz, 8 GB, and Windows 10 64bit. The POS Tagging application created using the C# programming language. The training process is carried out by changing the threshold to see its effect on the acquisition of rules, both on the lexical learner and contextual learner (GA). For testing, a trial is conducted to find out the accuracy of the training that has been done. For the calculation of the accuracy value, three types of calculations are used, namely the standard calculation without regard to OOV (Out of Vocabulary), the calculation by taking into account multiple POS and the calculation of accuracy by paying attention to the existence of OOV using equation(2) as in [24].

$$Accuracy\ OOV = \frac{Overall\ Accuracy}{(Known\ Word\ Acc/Unknown\ Word\ Acc)} \quad (2)$$

From the experimental results of the lexical learner for the 10 threshold produces 48 rules, the threshold of 20 to 40 has decreased the number of rules that is only 32 rules. Likewise for the 50 threshold produces the same rule as many as 13 rules. This shows that the smaller the threshold value, the more rules are produced. The greater the threshold value, the fewer rule will be produced. Result of Lexical Phase is shown below in Table I.

TABLE I. RESULT OF LEXICAL PHASE

T	Number of Rule	Accuracy (%)
10	54	85.81
20 30 40	33	85.81
50	13	85.12

The same thing happens in contextual learners with Brill Tagger, using threshold 2 produces 48 rules, threshold 3 produces 33 rules and threshold 4 produces 24 rules. Table II shows the best population results obtained during randomization to produce a Contextual Rule that produces the best accuracy on testing.

TABLE II. EXAMPLE OF CONTEXTUAL RULE WITH GA

No	Rules
1	SC VBT IN NN IN NN VBT
2	NN DT VBT NN NN NN CC
3	NN IN PRL VBT NN CC NN
4	NN VBI MD VBT ST Dummy Dummy
5	NN SC RB JJ VBT IN NN
6	IN JJ VBT NN RB VBT IN
7	NN VBI IN NN NN DT ST
8	NNP CP IN IN NN NN QT
9	NN JJ MD VBT NN SC VBT
10	SC VBT VBT NN IN NN VBT
11	DT NN SC VBI IN NN NN
12	NNP JJ IN NN RB VBT IN
13	CC WP NNG VBT NN SC NEG
14	NN IN PRL VBT NN CC NN
15	IN JJ VBT NN RB VBT IN

After conducting several contextual learner experiments with Brill Tagger and GA Brill Tagger by making threshold changes, the number of contextual rules is quite varied depending on the results of randomization. But the smaller the threshold, the more rules are obtained and the greater the threshold, the fewer rules are obtained.

TABLE III. RESULT OF CONTEXTUAL PHASE

T	Result of Lexical Phase	Accuracy (%)	
		Brill Tagger	***GA Brill Tagger***
2	85.81%	86.67	86.44
3		86.32	86.61
4		86.32	86.44

Table III shows the results of labeling with contextual rules using Genetic Algorithms. Accuracy has increased from lexical results, from an accuracy of 85.81% to 86.61%. Besides that, it is shown that the more rules that are produced (the smaller the accuracy), the better the accuracy tends to be. But in certain cases certain rules can justify the tag of a word and also give the wrong tag for other words. In the following table, we will show a cut of the results of the experiment using new data.

TABLE IV. RESULT WITH RESPECT OOV AND MULTIPLE POS

Manually Tagged Corpus	Brill Tagger	GA Brill Tagger
Dhinèng/SC	Dhinèng/SC	Dhinèng/SC
cara/NN	cara/NN	cara/NN
Mekkasán/NNP	Mekkasán/NNP	Mekkasán/NNP
kantos/VBT	kantos/VBT	kantos/VBT
Bhángkalan/NNP	Bhángkalan/NNP	Bhángkalan/NNP
kalèlès/NN	kalèlès/NNP	kalèlès/NN
èkaghábáy/VBT	èkaghábáy/VBT	èkaghábáy/VBT
dári/IN	dári/IN	dári/IN
kaju/NN	kaju/NN	kaju/NN
,/,	,/,	,/,
ghir-pèngghirra/NN	ghir-pèngghirra/NN	ghir-pèngghirra/NN
èokèr/VBT	èokèr/VBT	èokèr/VBT
,/,	,/,	,/,
èeccèt/VBT	èeccèt/VBT	èeccèt/VBT
ana-bárna/JJ	ana-bárna/JJ	ana-bárna/JJ
,/,	,/,	,/,
èparèngè/VBT	èparèngè/VBT	èparèngè/VBT
nèp-krennèp/VBT	nèp-krennèp/VBT	nèp-krennèp/NN
(/OP	(/OP	(/OP
ornamen/NN	ornamen/NN	ornamen/NN
)/CP	)/CP	)/CP
./ST	./ST	./ST
Bágiyán/NN	Bágiyán/NN	Bágiyán/NN
sè/SC	sè/SC	sè/SC
mennang/NN	mennang/JJ	mennang/NN
èkèrèm/VBT	èkèrèm/VBT	èkèrèm/VBT
ka/IN	ka/IN	ka/IN
Kerrap/NNP	Kerrap/NNP	Kerrap/NN
Gubeng/NNP	Gubeng/NNP	Gubeng/NNP
./ST	./ST	./ST

This stage depends on the rules generated in the genetic process where the resulting rules depend on the randomization process. Seen in table IV, the results with GA Brill Tagger for the word *nèp-krennèp* (glittering decoration) get the correct tag because of the rule "NN Prev1 / VBT" which means change the tag to NN if 1 tag was previously VBT, and for Brill Tagger results, the word *mennang* (win) gets the correct rule because the rule "NN JJ PREVWD sè" means that if the initial tag is NN and is located after the word sè then change the tag to JJ.

The experiment also be conducted by taking into account the existence of multiple POS. For example for the word *dháddi* which can have the tag as VBT in the sentence *èpateppa 'dháddi bhágus* (to be good) and as the SC in the *dháddi manabi sampèyan songkan entar ka dokter* (so if you are sick go to the doctor). From 2405 words and symbols in the corpus, there are several words have more than one POS. Table V shows example word with multiple POS.

TABLE V. EXAMPLE WORD WITH MULTIPLE POS

Word	POS 1	POS 2	POS 3
Kadháddhián	NN	VBI	
Lanjháng	NNP	JJ	
Kantos	VBT	RB	
Saè	JJ	SC	
Dháddi	VBT	SC	
Dálem	IN	NN	JJ
Dháddhi	VBT	SC	JJ

From a series of experiments with some changes in threshold values, the average accuracy obtained with Brill Tagger is 86.43% with the highest accuracy of 86.67%, while using GA Brill Tagger the average accuracy reaches 86.49% with the highest accuracy of 86.61%. Testing by considering multiple POS with Brill Tagger produces an average accuracy of 73.35% while testing using GA Brill Tagger produces an average accuracy of 90.93%. Testing with OOV produces an average accuracy of 67.22% with Brill Tagger and an accuracy of 64.58% with GA Brill Tagger. Table VI shows the experimental results related to OOV and Multiple POS.

TABLE VI. RESULT WITH RESPECT OOV AND MULTIPLE POS

Methode	OOV	Multiple POS
Brill Tagger	67.2%	73.4%
GA Brill Tagger	64.6%	90.9%

IV. CONCLUSION

Testing using Brill Tagger produces an average accuracy of 86.4% with the highest accuracy of 86.7% while testing using GA Brill Tagger produces an average accuracy of 86.5% with the highest accuracy of 86.6%. Testing by considering multiple POS with Brill Tagger produces an average accuracy of 73.4% while testing using GA Brill Tagger produces an average accuracy of 90.9%. Testing with OOV produces an average accuracy of 67.2% with Brill Tagger and an accuracy of 64.6% with GA Brill Tagger. This shows that the accuracy with GA Brill Tagger is better than Brill Tagger, especially if considering multiple POS. This is because GA Brill Tagger can generate rules for handling the existence of multiple POS more than pure Brill Tagger. For future work, the results of this study can be used to conduct other research on Bahasa Madura in the field of Natural Language Preprocessing such as Stemming, Question and Answering. This research can also be utilized for E-learning Bahasa Madura, and this very good because now Bahasa Madura has been abandoned by many Madurese people, especially among young people.

REFERENCES

[1] C. D. Manning, S. Hinrich, "Foundation of Statistical Natural Language Processing," Cambridge: MIT Press Textbook on statistical and probabilistic methods in NLP, 1999.

[2] E. R. Setyaningsih, "Penetapan Tagset dan Modifikasi Brill Tagger untuk Part-of Speech Bahasa Indonesia," Dinamika Teknologi, vol. 9 no.2, pp.37-42, 2017.

[3] K. Widhiyanti, A. Harjoko, "POS Tagging Bahasa Indonesia dengan HMM dan Rule Based," Jurnal Informatika, vol. 8 no.2, pp.151-167, 2012.

[4] F. Pisceldo, M. Adriani, R. Manurung, "Probabilistic Part Of Speech Tagging for Bahasa Indonesia," Third International MALINDO Workshop, 2009.

[5] V. Christanti, J. Pragantha, E. Purnamasari, "Implementasi Brill Tagger untuk memberikan POS-Tagging pada Dokumen Bahasa Indonesia," Jurnal Teknik dan Ilmu Komputer, 1(3), pp. 301–315, 2007.

[6] K. Kurniawan, A. F. Aji, "Toward a Standardized and More Accurate Indonesian Part-of-Speech Tagging," International Conference on Asian Language Processing (IALP), 10 September 2018, pp. 303–307, 2018.

[7] R. S.Yuwana, A. R. Yuliani, H. F. Pardede, "On Part of Speech Tagger for Indonesian Language," International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 1-2 Nopember 2017, pp. 369-372, 2017.

[8] F. M. Hasan, N. Uzzaman, M. Khan, "Comparison of different POS Tagging Techniques (N-Gram, HMM and Brill's tagger) for Bangla," Advances and Innovations in Systems, Computing Sciences and Software Engineering, pp.121-126, 2007.

[9] E. Brill, "A simple rule-based part of speech tagger," Proc. third Conf. Appl. Nat. Lang. Process, pp. 152, 1992.

[10] N. P. M. Sriyati, "Part-Of-Speech Tagging Untuk Dokumen Bahasa Bali Menggunakan Algoritma Brill Tagger," Fakultas Matematika dan Ilmu Pengetahuan Alam. Tugas Akhir. Universitas Udayana, 2016.

[11] K. Singh, "Part-of-Speech Tagging using Genetic Algorithms," International Journal of Simulation - Systems, Science & Technology, 16(6), pp.111-117, 2015.

[12] A. P. Silva, A. Silva, I. Rodrigues, "A New Approach to the POS Tagging Problem Using Evolutionary Computation," Proceedings of Recent Advances in Natural Language Processing, 7-13 September 2013, pp. 619-625, 2013.

[13] K. Nongmeikapam, S. Bandyopadhyay, "Genetic Algorithm (GA) Implementation for Feature Selection in Manipuri POS Tagging," Proceedings of the 13th International Conference on Natural Language Processing, Desember 2016, pp. 267–274, 2016.

[14] Wilson, Garnett, Malcolm Heywood, "Use of a Genetic Algorithm in Brill's Transformation-Based Part-of-Speech Tagger," Faculty of Computer Science Dalhousie University, 2005.

[15] J. Bjerva, "Genetic Algorithms in the Brill Tagger-Moving towards language independence," Department of Linguistics. Thesis. Stockholm University, 2013.

[16] A. Halim, "Politik Bahasa Nasional 1 dan 2," Jakarta: Aneka Ilmu, 1976.

[17] N. P. Dewi, Ubaidi, "Lexical Rule dan Pengaruh Penggunaan Lexicon Pada Pos Tagging Bahasa Madura," Jurnal Matrik, 18(1) pp.69-70, 2018.

[18] N. P. Dewi, Ubaidi, "POS Tagging Bahasa Madura dengan Menggunakan Algoritma Brill Tagger," Jurnal Teknologi Informasi dan Ilmu Komputer, unpublished.

[19] A. G. Ayana, "Improving Brill's Tagger Lexical and Transformation Rule for Afaan Oromo Language," PeerJ PrePrints, pp.1-11, 2015.

[20] M. Mitchell, "An Introduction to Genetic Algorithms," Cambridge, MA: MIT Press, 1996.

[21] A. E. Eiben, "Genetic algorithms with multi-parent recombination," PPSN III: Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature: 78–87, 1994.

[22] D. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning," Reading, MA: Addison-Wesley Professional, 1989.

[23] A. Dinakaramani, F. Rashel, A. Luthfi, R. Manurung, "Designing an Indonesian Part of speech Tagset and Manually Tagged Indonesian Corpus," International Conference on Asian Language Processing (IALP), 20-22 Oktober 2014, pp. 66-69, 2014.

[24] Wicaksono, A. Farizki, A. Purwarianti, "HMM Based Part-of-Speech Tagger for Bahasa Indonesia," On Proceedings of 4th International MALINDO (Malay and Indonesian Language) Workshop, 2nd August 2010